## PATENT ABSTRACTS OF JAPAN

(11)Publication number:

09-179694

(43)Date of publication of application: 11.07.1997

(51)Int.CI.

G06F 3/06

(21)Application number: 07-338611

(71)Applicant:

MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing:

26.12.1995

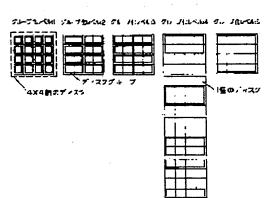
(72)Inventor:

MORI YASUHIRO

## (54) FILE SERVER

#### (57)Abstract:

PROBLEM TO BE SOLVED: To provide a file server capable of serving continuous media data having respectively different bit rates. SOLUTION: In the file server provided with an external storage device constituted of N disks, the number of grouped disks is changed in each grouping level L to organize disk groups. Continuos media data are divided into plural blocks, grouping levels corresponding to respective blocks are determined in accordance with their bit rates and the blocks are allocated to the groups with the grouping levels. The space of each disk is divided into byte units, contents existing on the same address of all the disks are transversally united and an area consisting of S × N bytes is constituted as a package and used for the allocation unit of storage space to the continuous media data.



### **LEGAL STATUS**

[Date of request for examination]

15.10.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3570050

[Date of registration]

02.07.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of

rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

技術表示箇所

(19)日本国特許庁(JP)

# (12) 公開特許公報(A)

(11)特許出願公開番号

# 特開平9-179694

(43)公開日 平成9年(1997)7月11日

(51) Int.Cl.<sup>6</sup>

G06F 3/06

識別記号 540

庁内黎理番号

FΙ G06F

540

審査請求 未請求 請求項の数6 〇L (全 9 頁)

(21)出願番号

(22)出願日

特顯平7-338611

平成7年(1995)12月26日

(71)出願人 000005821

3/06

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 森 康浩

大阪府門真市大字門真1006番地 松下電器

産業株式会社内

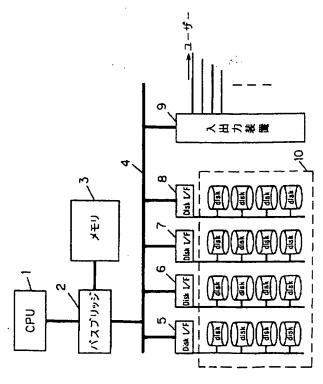
(74)代理人 弁理士 滝本 智之 (外1名)

## (54) 【発明の名称】 ファイルサーパ

## (57) 【要約】

【課題】 ビットレートの異なる連続メディアデータを サービス可能なファイルサーバを提供する。

【解決手段】 N台のディスクから構成される外部記憶 装置を備えたファイルサーバにおいて、グループ化レベ ル:L毎にグループ化ディスク数:Dを変えてディスク グループを編成する。連続メディアデータはブロックに 分割し、そのピットレートに応じて対応するグループ化 レベルを決定し、そのグループ化レベルのグループに対 してプロックを割り当てる。また、ディスクの空間をS10 バイト単位に分割するとともに、全てのディスク上で同 ーアドレスにあるものを横断的に一元化し、S×Nバイ トの領域をパッケージとして連続メディアデータへの記 **億空間の割当ての単位とする。** 



2

## 【特許請求の範囲】

【請求項1】N台(Nは自然数)のディスクから構成される外部記憶装置を備えたファイルサーバであって、

自然数: L に対して、D = f (L) である関数: f () があって、

前記N台のディスクに対して、それぞれがD個のディスクから構成されるディスクグループをM個編成することをグループ化レベル:Lにおけるグループ化と定義する時、

この外部記憶装置にアクセスされるデータのアクセス単10位に応じて、Lを変化させることを特徴とするファイルサーバ。

【請求項2】 M個のディスクグループのそれぞれを一定時間毎に巡回的にアクセスすることによって外部記憶装置に対する一定レートの入出力を保証し、Lを変化させることによって様々なレートの入出力を行なうことを特徴とする請求項1記載のファイルサーバ。

【請求項3】時間的な属性がビットレートによって定義される連続メディアデータをブロック単位に分割してN台のディスクに分散記憶する場合に、

前記ピットレートで一定時間に入出力されるデータをアクセス単位とし、このアクセス単位に応じたグループ化レベル: L において編成されたディスクグループ群に前記ブロックを分散配置することを特徴とする請求項1記載のファイルサーバ。

【請求項4】外部記憶装置の各々のディスクの記憶空間をSバイト(Sは任意の自然数)単位のセグメントに区切り、前記外部記憶装置を構成するN個のディスク間で同一アドレスにあるS×Nバイトの集合体をパッケージとし、 30

このパッケージを連続メディアデータに対する前記外部 記憶装置の記憶空間の割り当ての単位とすることを特徴 とするファイルサーバ。

連続メディアデータのデータブロックが外部記憶装置の 40 アドレス (sn, offset, dn) に対してdnを +1し、

dnが前記Nを越えた場合にはdnをリセットして、off set にブロックサイズを加算するステップをoff fset がセグメントサイズを越えるまで繰り返すようにして、パッケージ内に前記ブロックを配置することを特徴とする請求項4 記載のファイルサーバ。

【請求項6】外部記憶装置に対してグループ化レベル: Lの異なる複数の連続メディアデータがアクセスされ、 それぞれの連続メディアデータが対応するグループ化レ50 ベルのディスクグループからブロックを読み出す際のディスクグループの番号の遷移速度と遷移順序が等しいことを特徴とするファイルサーバ。

## 【発明の詳細な説明】

[0001]

【発明の属する技術分野】本発明はデータの正当性に加えて、再生時の時間的な正当性を要求する連続メディアデータを扱うためのファイルサーバに関する。

[0002]

【従来の技術】従来のUNIXにおけるファイルシステムでは、ディスクの記憶空間は物理的に連続したセクタの集合で、固定長の物理プロック単位で管理され、この物理プロックがディスクからのread/write、ならびにファイルへの割当ての単位となっている。

【0004】iノードを利用したディスクアクセスの場合、ファイルサイズが小さい場合は直接アドレスで変換可能であるが、ファイルサイズが大きくなると間接アドレスを多用することになる。一般に、従来のUNIXで扱ってきたデータはテキストデータに代表されるデータであり、そのサイズも大半が10Kバイトを越えないといった研究もなされている。このため、先述したようなメカニズムによるファイルへのアクセス方式でもそのオーバーヘッドは問題にならなかった。

【0005】また、UNIXではディスクアクセスのオーバーヘッドを削減するために、パッファキャッシュが実装されている。これはデータアクセスに対する空間的、時間的な局所性を生かしたもので、一旦ディスクからフェッチした物理ブロックはメモリ上にキープしておき、これを再利用するためのメカニズムであり、これによってディスクへのアクセス頻度を削減してデータアクセスのオーバーヘッドを削減できる。

【0006】さらに、テキストデータはデータの正当性のみを要求し、時間的な正当性を要求しない。したがって、特定のディスクにアクセスが集中して、読み出し要求に対する応答速度が遅くなっても特に大きな問題ではなかった。

4

【発明が解決しようとする課題】しかしながら、連続メディアデータは従来のテキストデータと違って、データの正当性に加えて、再生時の時間的な正当性を要求するデータであり、その特徴として、

・データのサイズが桁違いに大きい。

【0008】・データのアクセスに対する時間的・空間 的な局所性がない。

・データのアクセスに対する連続性、あるいは予測可能 性がある

といった点を挙げることができ、従来のテキストデータ のようなデータをターゲットとしたファイルサーバで扱 うには問題が大きい。

【0009】例えば、データのサイズが桁違いに大きいので、従来のiノードによるアドレス変換では最終的な物理ブロックへのポインタを得るまでに数回間接アドレスを手繰る必要があり、その度にディスクアクセスが必要である。

【0010】また、データのアクセスに対しては、時間的・空間的な局所性がないということは、つまり、先の20サイクルで読みだした物理ブロックは次のサイクルにおいては必要でない可能性が非常に高いということである。逆にアクセスに対する連続性・予測可能性が高いので、これを生かすためにバッファキャッシュのようなメカニズムよりもデータをプリフェッチするメカニズムの方が有効であることになる。データをプリフェッチするためにはディスク上でリニアに配置された方が都合が良いが、iノードによる変換方式では、あるファイルに割り当てられる物理ブロックはディスク上の記憶空間上で離散的に置かれることになり、先述したデータのアクセ30スに対する連続性、予測可能性を生かせない。

【0011】本発明は上記のような問題を鑑みてなされたものであり、連続メディアデータの特徴を生かしたファイルサーバを提供することを目的としている。

### [0012]

【課題を解決するための手段】この目的を達成するため、第1の発明においては、N台(Nは自然数)のディスクから構成される外部記憶装置を備えたファイルサーバであって、自然数:Lに対して、D=f(L)である関数:f()があって、前記N台のディスクに対して、40それぞれがD個のディスクから構成されるディスクグループをM個編成することをグループ化レベル:Lにおけるグループ化と定義する時、この外部記憶装置にアクセスされるデータのアクセス単位に応じて、Lを変化させることを特徴とするファイルサーバとしている。

【0013】第2の発明においては、第1の発明の構成に加え、M個のディスクグループのそれぞれを一定時間毎に巡回的にアクセスすることによって外部記憶装置に対する一定レートの入出力を保証し、Lを変化させることによって様々なレートの入出力を行なうことを特徴と50

するファイルサーバとしている。

10

【0014】第3の発明においては、第1の発明のファイルサーバにおいて、時間的な属性がビットレートによって定義される連続メディアデータをブロック単位に分割してN台のディスクに分散記憶する場合に、前記ビットレートで一定時間に入出力されるデータをアクセス単位とし、このアクセス単位に応じたグループ化レベル: Lにおいて編成されたディスクグループ群に前記ブロックを分散配置することを特徴とするファイルサーバとしている。

【0015】第4の発明においては、外部記憶装置の各々のディスクの記憶空間をSバイト(Sは任意の自然数)単位のセグメントに区切り、前記外部記憶装置を構成するN個のディスク間で同一アドレスにあるS×Nバイトの集合体をパッケージとし、このパッケージを連続メディアデータに対する前記外部記憶装置の記憶空間の割り当ての単位とすることを特徴とするファイルサーバとしている。

【0017】第6の発明においては、外部記憶装置に対してグループ化レベル: Lの異なる複数の連続メディアデータがアクセスされ、それぞれの連続メディアデータが対応するグループ化レベルのディスクグループからブロックを読み出す際のディスクグループの番号の遷移速度と遷移順序が等しいことを特徴とするファイルサーバとしている。

【0018】上記構成によって、第1の発明において、N台のディスクはグループ化レベル: L毎に段階的にグループ化するディスクの数: Dを変えてグループ化することによって、アクセスされるデータのアクセス単位に対応し、Lを変化させることによって様々なアクセス単位に対応可能としている。

【0019】また、第2の発明において、第1の発明の作用に加えて、グループを構成するディスクを並列にアクセスすることによって仮想的に高速/大容量の1個のディスクとして一定レートの入出力を保証し、また、Lを変化させることによってグループ化するディスクの数

【0020】また、第3の発明においては、連続メディアデータを分割したブロックの割り当て先として、ビットレートによって決まるグループ化レベルのグループとすることで、ビットレートに応じたディスクのアクセスレートを確保し、また、ディスクの増設時のデータのスケーラビリティーを保持できる。

【0021】また、第4の発明において、S×Nバイトの空間をパッケージと呼び、これをファイルに対する記10 憶空間の割当の単位とすることによって、ディスク空間の管理テーブルを小さくし、また、ディスク空間の管理単位とディスク空間へのread/writeの単位を分割できるとしている。

【0022】また、第5の発明において、請求項4の作用に加えて、パッケージのアドレス指定をセグメント番号:snと、ディスクに付与されたディスク番号:dnと、セグメント内でのオフセット値:offsetの組(sn,offset,dn)によって定義し、この組の中でdn,offsetの順でより右にある変数をよ20り頻繁に変化させることによって、パッケージ内で論理的に連続してデータを配置することが可能となるとしている。

【0023】また、第6の発明において、グループ化レベルのディスクグループからブロックを読み出す際のディスクグループの番号の遷移速度と遷移順序を等しくすることによって、特定のディスクに対するアクセスの集中を避けることが可能としている。

[0024]

【発明の実施の形態】本発明の一実施の形態のファイル 30 サーバについて、図面を参照しながら説明する。

【0025】 (実施の形態1) 図1は本発明の第1の実

施の形態のファイルサーバのシステム構成を示すプロック図である。CPU1とメモリ3とこれらと拡張バス4を接続するバスプリッジ2から構成される一般的なパーソナルコンピュータ(PC)の構成であって、拡張バス4にはn個のディスクインターフェース(5~8)が接続され、そして、それぞれのディスクインターフェースにm台のディスクが接続されて、合計パーn×m台のディスクによって外部記憶装置10が構成されている。また、ユーザーへのデータの配信と外部記憶装置10への

6

【0026】外部記憶装置10に格納された連続メディアデータはCPU1の制御によって読み出されて、一時的にメモリ上に置かれ、あるいは直接に、入出力装置9に送り込まれ、その後ユーザーに配信される。

データの格納のために拡張バス4に入出力装置9を接続

【0027】図2は本実施の形態のファイルサーバにおけるにおけるディスクグループの編成を示すブロック図である。外部記憶装置はN台のディスクから構成されるが、ここでは説明のためにn=m=4とし、合計16台のディスクによって構成されているものとしている。また、 $L \rightarrow D$ の対応は(数1)に示す関数によって定義しているが、必ずしもこの数式に拘る必要はない。図2中で一つのマスが1個のディスクを表し、太線枠で囲まれたものが一つのディスクグループである。

[0028]

している。

【数1】

 $D = f(L) = 2^{(L-1)}$ 

【0029】グループ化レベル: Lと各々のグループ化レベルにおけるグループ化ディスク数: Dを(表1)に示す。

[0030]

【表1】

| グループ化レベル:L | グループ化ディスク数:D |
|------------|--------------|
| 1          | 1            |
| 2          | 2            |
| 3          | 4            |
| 4          | 8            |
| 5          | 16           |

【0031】(表1)ではグループ化レベルが2の場合、16個のディスクを2個づつグループ化し、合計8個のグループができることを示している。

【0032】グループ化されたディスクは仮想的に一個のディスクとみなし、データのアクセス単位によってグ50

ループ化レベルを変化させて、様々なアクセス単位のデータに対応可能とする。

【0033】(実施の形態2)図3は本発明の第2の実施の形態を示すプロック図である。図3は4×4個のディスクに対して、グループ化レベル3でグループ化する

場合について書いている。グループを構成するディスクは並列にアクセスする。これによって、(一つのディスクからの読みだしレート)×(ディスク数)の読みだしレートを確保することが可能となる。また、グループを逐次的に、かつ全体として周期的にアクセスすることにより特定のディスクへのアクセスの偏りを避けることができる。

【0034】(実施の形態3)図4は本発明の第3の実施の形態を示すブロック図であって、連続メディアデー

タを分割し、これをディスクグループへ割当てる方法を示したものである。まず、連続メディアデータはブロック単位に分割し、先頭から順にプロック番号を振る。次に、連続メディアデータのビットレートによって割り当て先のディスクのグループ化レベルを決定する。 (表2) はその対応についての一例を示したものである。【0035】

【表2】

| ピットレート(Mbps) | グループ化レベル:L | グループの数:GN |  |
|--------------|------------|-----------|--|
| 1 ~ 7        | 1          | 16        |  |
| 8 ~ 15       | 2          | 8         |  |
| 16 ~ 31      | 3          | 4         |  |
| 32 ~ 63      | 4          | 4         |  |
| 64 ~ 100     | 5          | 1         |  |

【0036】対応するグループ化レベルが決定したら、そのグループ化レベルにおけるグループ数:GNで、ブロック番号を割り、その剰余によってブロックをクラス分けして割り当て先のグループを決定する。

【0037】プロックの割り当て先をビットレートによって決まるグループとすることによるメリットの一つにディスクを拡張した場合のデータの互換性が挙げられ 30 る。4×4のディスク構成から1行分のディスクを増設して、5×4のディスク構成とした場合、グループの編成が変わるだけであって、データのフォーマットそのものには何ら影響がない。したがって、ディスクの拡張が行なわれた場合には、編成の変更に合わせてデータプロックを移動させてやれば良いのであって、データそのものを作り直す必要がない。

【0038】(実施の形態4)図5は本発明の第4の実施の形態を示すブロック図であって、ディスクの記憶空間をパッケージ単位に分割し、このパッケージを連続メ40ディアデータへの割り当ての単位とすることを示したものである。

【0039】図5に示すように、N台のディスクから構成される外部記憶装置において、それぞれのディスクの記憶容量をMバイト、ディスク上で物理的に連続した領域をセグメント、そのサイズをSバイトとするとき、それぞれのディスク上の記憶空間をM/S個のセグメントに分割する。さらにディスク間で同一のアドレスにあるセグメントを横断的に一くくりにしたS×Nバイトの領域をパッケージとする。連続メディアデータを外部記憶50

装置に書き込む際にはこのパッケージをファイルに対するディスク空間の割当ての単位とし、read/writeの単位である物理ブロックとは切り離す。

【0040】ディスク空間の管理単位をパッケージとすることは、連続メディアデータの記憶媒体の容量がギガオーダーになることが通常でありがちなファイルサーバにとって、管理テーブルを小さくでき、システム領域として必要なメモリの節約に継る。また、様々なビットレートの連続メディアデータをターゲットとするファイルサーバにとって、一定時間にアクセスされるデータのアクセス単位はビットレートによって定義する方が都合いが、ディスク空間の割り当ての単位もデータによって変化するとディスク空間のフラグメンテーションの問題が残る。したがって、read/writeの単位とディスク空間の割当ての単位を分離することはフラグメンテーションの問題を解決し、かつ、様々なピットレートのデータを扱う上で最も都合の良い選択子であるといえる。

【0041】(実施の形態5)図6は本発明の第5の実施の形態を示すフローチャート図であって、連続メディアデータを記憶するために割り当てられたパッケージに対して、ブロックを書き込む際のアルゴリズムを示している。

【0042】パッケージには各々のディスク内で一元的に付与されたセグメント番号: $sn \ge$ 、ディスクに一元的に付与されたディスク番号: $dn \ge$ セグメントの先頭アドレスからのオフセット値:offsetによってア

ドレス (pn, offset, dn) が振られていて、このアドレスが指し示すポイントからデータのread /writeを行なう。

【0043】まず、最初のブロックはアドレス(sn, of fset, dn)の示す領域に書き込む。次にブロック番号を+1する。アドレスはdnを+1し、dnが Nを越えない場合はディスク番号だけが違う領域にブロックを書き込む。このとき、dnがNを越えた場合は、一旦dnをクリアし、of fsetにブロックサイズを加算する。

【0044】そして、offsetがセグメントサイズを越えない場合は、(sn, offset, dn)の示す領域にブロックを書き込む。また、このとき、offsetがセグメントサイズを越えた場合は、パッケージがフルの状態であり、新しいパッケージを獲得してブロックの書き込みを続ける。

【0045】本方法によって、ブロックはパッケージ上で論理的に連続して配置されることになる。したがって、次のサイクルで必要なブロックのアドレスを得ることが比較的容易に可能であるので、連続メディアデータ 20のアクセスの連続性・予測可能性を生かしやすい。

【0046】(実施の形態7)図7は各々のグループ化レベルのディスクグループ群中でのグループの読み出し順序を定義したものである。図中の数字はそのグループ化レベルにおけるグループの読み出し順序を表し、サービスサイクル: tにおいて、1のグループを読みだした場合は、t+1においては2のグループを読み出すことを示している。数字は周期的に使用する。レベル5においては、毎サイクル全てのディスクが並列アクセスされることになる。

【0047】図8は図7で定義された順序にしたがっ て、2つの異なるグループ化レベルに対応するデータを 読み出す場合のディスクアクセスの順序を示すブロック 図である。図7から明らかなようにグループ化レベルの 異なる連続メディアデータを読み出す場合でも、ディス ク間の遷移速度が等価である。これは、つまり、t=0 で2つのデータを読み出し可能であるならば、あらゆる サイクルでデータの読み出しが可能であることを意味し ている。図7の示すルールに従わない場合、最悪の場 合、グループ化レベルの異なる2つ以上のデータを読み40 出すと、グループ化レベル毎のディスクアクセス周期の 最小公倍数毎にアクセスが特定のディスクに集中する。 しかしながら、図7に示した順序によって、外部記憶装 置をアクセスする場合、あるディスクグループのあるデ ィスクからのデータ読み出し量は次のサイクルに於て隣 のディスクグループ内のディスクに継承される。したが って、あるディスクから複数のデータを読み出し可能で あるならば、サービスサイクルが変わっても常に読み出 し可能となる。

【0048】本発明の有利な点は、ディスクアクセスに50

関して特定の周期で特定のディスクに対するアクセスの 集中を回避できるため、ある時間においてすべてのディ スクから要求されたデータを一定時間内に読みだし可能 であるならば、それ以後いかなる時間においてもすべて のディスクから要求されたデータが読みだし可能である ということである。

【0049】連続メディアデータを扱うファイルサーバにおいては、一定時間内に要求されたデータを外部記憶装置から読み出すことが重要であり、それが不可能であるということは、データの再生時に映像が乱れるなどのエラーを引き起こすことに継り、致命的である。

### [0050]

10

【発明の効果】本願によって開示される発明のうち代表的なものによって得られる効果を簡単に説明すれば、下記の通りである。

【0051】すなわち、N個のディスクから構成される外部記憶装置をグループ化レベル毎にグループ化するディスクの数を変えてグループ単位に編成し、このグループ化レベルと連続メディアデータのピットレートを分割したブロックををそのグループ化レベルのグループに割当て、データの読み出し時にこのグループ内のディスクを並列にアクセスとし時にこのグループ中のディスクを進列としたが可能である。また、グループ化レベル毎にディスクグループ単位でブロックの読み出しを行うことで、異なるどが可能である。また、グループ化レベル毎にディスクグループ単位でブロックの読み出しを行うことで、異なるどが写してある。また、グループのであっても、ディスクの選移速度が等により、特定のディスクへのアクセスの集中が回避できる。

【0052】さらに、パッケージ単位で記憶空間を割り当て、その中で連続的にブロックを配置することで、連続メディアデータの連続性・予測可能性を生かすことが可能となる。さらに、記憶空間の割当ての単位をread/writeの単位と切り離すことで管理テーブルのサイズを小さくするとともに、ディスク空間の割り当てにおけるフラグメンテーションを解決する上で重要なポイントとなる。

## 【図面の簡単な説明】

【図1】本発明のファイルサーバのシステム構成を示す ブロック図

【図2】本発明の第1の実施の形態を示すブロック図

【図3】本発明の第2の実施の形態を示すブロック図

【図4】本発明の第3の実施の形態を示すプロック図

【図5】本発明の第4の実施の形態を示すブロック図

【図6】本発明の第5の実施の形態を示すブロック図

【図7】本発明の第6の実施の形態を示すブロック図

【図8】同実施の形態における動作を示す週移図 【符号の説明】

1 CPU

2 メモリ

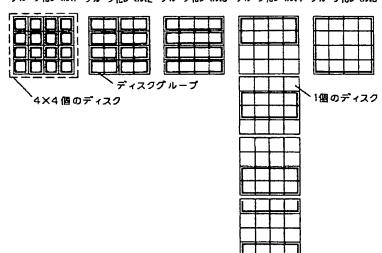
- 3 バスブリッジ
- 4 拡張バス
- 5 ディスクインターフェース
- 6 ディスクインターフェース

- 7 ディスクインターフェース
- 8 ディスクインターフェース
- 9 入出力装置
- 10 外部記憶装置

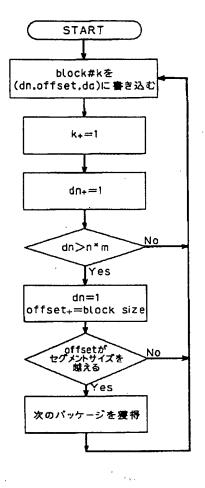
【図1】

[図2]

グループ化レベル:1 グループ化レベル:2 グループ化レベル:3 グループ化レベル:4 グループ化レベル:5

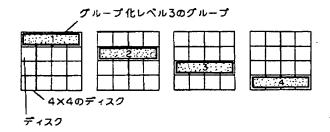


【図6】

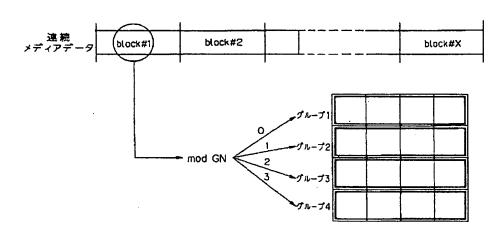


4

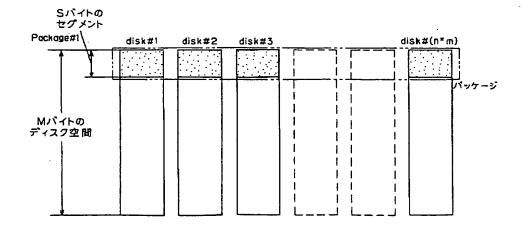
[図3]



[図4]



【図5】



【図7】

| グループ化レベル:1                                      | グループ化レベル:2               | グループ化レベル:3  | グループ化レベル:4 | グループ化レベル:5 |
|---|--------------------------|-------------|------------|------------|
| 1 9 5 13<br>2 10 6 14<br>3 11 7 15<br>4 12 8 16 | 1 5<br>2 6<br>3 7<br>4 8 | 2<br>3<br>4 | 1          | 1          |
|   |                          |             | 2          |            |
|   |                          |             | 3          |            |
|   |                          |             | 4          |            |

【図8】

